6. LOW-INFERENCE TEACHING BEHAVIORS AND COLLEGE TEACHING EFFECTIVENESS: RECENT DEVELOPMENTS AND CONTROVERSIES

Harry G. Murray
University of Western Ontario
murray@uwo.ca

Key Words: Teacher characteristics, low-inference teaching behaviors, student ratings of instruction, effective college teaching

INTRODUCTION

Research on teacher effectiveness in higher education attempts to specify characteristics of teachers that contribute to the cognitive or affective development of students (Murray, 1991). It is assumed that knowledge of teacher characteristics contributing to effective teaching will lead both to a better theoretical understanding of teaching and to the development of improved programs of faculty selection, faculty evaluation, and faculty development.

One of the first and most important problems that must be faced in teacher effectiveness research is that of criterion measurement. Measures of teaching effectiveness used in higher education studies to date have included: (1) student learning—for example, mean student performance on a common final examination in a multiple-section course; (2) student motivation for further learning—for example, frequency of students enrolling in advanced courses in the teacher’s area of study; and (3) formal student ratings of instructional quality. Student instructional ratings, the most frequently used criterion measure in teacher effectiveness

This paper is based on an invited presentation to the Special Interest Group on Faculty Evaluation and Development, American Educational Research Association, at the 1997 AERA Annual Meeting in Chicago (Murray, 1997).

© 2007 Springer.
research, provide both a direct measure of student satisfaction with
instruction and an indirect or "proxy" measure of outcome variables
such as student learning and student motivation. Evidence that student
ratings are suitable or appropriate as a direct or indirect measure of
teacher effectiveness includes the following: (1) high retest and in-
trater reliability; (2) moderate to high agreement with evaluations of
the same instructors by other independent judges; (3) generally weak
correlation with extraneous factors such as class size, strictness of grading,
and course level; and (4) significant correlation with more objective
indicators of teaching effectiveness, such as student learning and student

Several different types of research design have been employed in
teacher effectiveness research in higher education, including
survey research, case studies, ethnography, classroom observation, and
laboratory experimentation. The present review deals mainly with
research using observational and experimental designs. In observational
research, teachers are studied under natural conditions (usually in the
classroom) with no manipulation or control of variables, and observed
teacher characteristics are analyzed in relation to outcome measures
such as student exam performance or student instructional ratings.
Observational findings tend to be high in external validity but low in
internal validity. In other words, research findings are generalizable to
actual classrooms, but it is difficult to determine whether corre-
lations found between teacher characteristics and outcome measures
represent true cause-effect relationships. In experimental research,
teachers are studied under laboratory or field conditions where one
or more instructional variables are systematically manipulated by the
investigator, with all other variables controlled or held constant. Under
experimental conditions, it is possible to infer cause-effect relationships
between teacher characteristics and measures of effectiveness (high
internal validity), but the contrived artificiality of laboratory experi-
ments may limit generalizability of results to real classrooms (low
external validity). The ideal situation, of course, is for research findings
to be replicated in both observational and experimental designs, so that
the strengths of one type of design compensate for the weaknesses of
the other, and the overall credibility of findings is maximized.

Two distinct types of instructional variables have been studied in
teacher effectiveness research in higher education: high-inference and
low-inference (Rosenshine and Furst, 1971). High-inference teacher
characteristics are global, abstract traits such as "explains clearly" or
"has good rapport," while low-inference characteristics are specific,
concrete teaching behaviors, such as "signals the transition from one
topic to the next" and "addresses individual students by name," that
can be recorded with very little inference or judgement on the part of
a classroom observer. Rosenshine and Furst's high-inference vs. low-
inference dichotomy relates to Feldman's (1998) discussion of amount
of inferring done by students in evaluating teachers as one of the dimen-
sions contributing to the subjectivity-objectivity of student instruc-
tional ratings. Table 1 gives further examples of high-inference teacher
traits and corresponding low-inference teacher behaviors. Although
knowledge of both low-inference and high-inference characteristics
is needed for a full understanding of teaching effectiveness, it can be
argued that there are some definite advantages in focusing on specific,
low-inference teaching behaviors. For one thing, such behaviors are
relatively easy to manipulate or record for research purposes, and
researchers are more likely to use consistent operational definitions
of teaching when they are based on specific, concrete behaviors.
Second, low-inference behaviors are valuable in teaching improvement
programs because they provide specific, concrete examples of effective
teaching that are easier to acquire or modify than high-inference char-
acteristics such as "clear" or "coherent." Finally, low-inference classroom
teaching behaviors represent the "leading edge" of teaching, the point
of direct contact between teacher and student, and thus (it is hoped)
are more likely to have a direct impact on student development than
high-inference teacher characteristics such as subject knowledge,
goals, planning. I don't mean to imply that we should ignore these
high-inference characteristics in teaching, but it seems to me that good
planning and good intentions on the part of the teacher will go for
nothing unless these plans and intentions are translated into specific,
effective classroom behaviors.

<table>
<thead>
<tr>
<th>Table 1: High-Inference vs. Low-Inference Teacher Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High-Inference</strong></td>
</tr>
<tr>
<td>Organization</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Enthusiasm</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
This chapter provides a brief review of early research on low-inference teaching behaviors in relation to student instructional ratings, then a more detailed review of recent low-inference studies that have dealt with the following issues or questions: (1) Are low-inference teaching behaviors related to outcome measures other than student ratings, and if so, to what measures? (2) Is there a cause-effect relationship between low-inference teaching behaviors and measures of teaching effectiveness? (3) What are the cognitive or affective processes that underlie or mediate the relationship between low-inference teaching behaviors and student outcome measures? (4) Is the relationship between low-inference behaviors and student instructional ratings consistent across different situations or contexts? and (5) Can research on low-inference teaching behaviors be successfully applied to programs for improvement of teaching?

EARLY RESEARCH

Early research by myself and others showed that there is indeed a clear relationship between low-inference classroom teaching behaviors and student ratings of overall teaching effectiveness (e.g., Cranton and Hillgartner, 1981; Mintzes, 1979; Murray, 1983a, 1983b, 1985; Tom and Cushman, 1975). My own studies were designed to systematically observe and compare the teaching behaviors of instructors receiving low, medium, and high ratings from students, and thereby to determine specifically what it is that highly rated teachers actually do in the classroom, or less highly rated teachers fail to do. These studies involved sending trained observers into regular classes taught by participating faculty members. Typically each of 40 to 50 instructors was observed in three separate randomly selected one-hour class periods by each of 6 to 8 observers. The observers were undergraduate students who were paid, given preliminary training in recording teaching behaviors, and told to be as unobtrusive as possible in the classroom. The style of teaching was lecture or lecture-discussion, with a minimum class size of 30. Instructors gave informed consent for participation, but did not know the exact dates on which classroom observation would occur. Following their 3 hours of observation of a given instructor, observers completed an instrument known as the Teacher Behaviors Inventory (TBI), illustrated in part in Table 2.12 which required ratings of the frequency of occurrence of each of 50 to 100 low-inference teaching behaviors on a 5-point scale ranging from 1 (almost never) to 5 (almost always). Ratings were averaged across observers to obtain mean

| Table 2: Teacher Behaviors Inventory |
| Research Version |
|---|---|---|---|---|---|
| Instructions to observer |
| A: Almost Never | B: Rarely | C: Sometimes | D: Often | E: Almost Always |
| 1. gives several examples of each concept | 2. uses concrete, everyday examples to explain concepts and principles | 3. fails to define new or unfamiliar terms | 4. uses graphs and diagrams to facilitate explanation | 5. repeats difficult ideas several times | 6. stresses most important points by pausing, speaking slowly, raising voice, etc. | 7. suggests ways of memorizing complicated ideas | 8. writes key terms on blackboard or overhead screen |

frequency estimates for each teaching behavior and each instructor, then statistical analyses were run to identify specific teaching behaviors that correlated significantly with student ratings of overall teaching effectiveness, measured either by a single, global item or by the average of all items on a formal end-of-term teaching evaluation form. One advantage of the research design used in these studies is that classroom teaching behaviors are assessed in a relatively nonobtrusive or nonreactive way. Another advantage is that independent and dependent variables are measured in procedurally independent ways (i.e., low-inference teaching behaviors by classroom observers, overall teaching effectiveness by student raters), thus minimizing the possibility of spurious correlations between teaching behaviors and overall effectiveness ratings due to "halo effect" or "implicit personality theory."
Table 3: Factor Loadings, Interrater Reliabilities, and Correlations with Student Rating of Overall Teaching Effectiveness Rating for 27 Low-Inference Teaching Behaviors

<table>
<thead>
<tr>
<th>Teaching Behavior</th>
<th>Factor Loading</th>
<th>Interrater Reliability</th>
<th>Correlation with Student Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>.57</td>
<td>.76</td>
<td>.47*</td>
</tr>
<tr>
<td>uses concrete examples</td>
<td>.49</td>
<td>.78</td>
<td>.61*</td>
</tr>
<tr>
<td>stresses most important points</td>
<td>.64</td>
<td>.66</td>
<td>.30*</td>
</tr>
<tr>
<td>repeats difficult ideas</td>
<td>.69</td>
<td>.84</td>
<td>.42*</td>
</tr>
<tr>
<td>Expressiveness</td>
<td>.70</td>
<td>.89</td>
<td>.38*</td>
</tr>
<tr>
<td>shows facial expressions</td>
<td>.76</td>
<td>.78</td>
<td>.63*</td>
</tr>
<tr>
<td>gestures with hands and arms</td>
<td>.51</td>
<td>.92</td>
<td>.36*</td>
</tr>
<tr>
<td>speaks expressively or “dramatically”</td>
<td>.82</td>
<td>.86</td>
<td>.26*</td>
</tr>
<tr>
<td>Interactions</td>
<td>.54</td>
<td>.77</td>
<td>.36*</td>
</tr>
<tr>
<td>addresses individual students by name</td>
<td>.80</td>
<td>.81</td>
<td>.21*</td>
</tr>
<tr>
<td>asks questions as a whole</td>
<td>.57</td>
<td>.66</td>
<td>.51*</td>
</tr>
<tr>
<td>praises students for good ideas</td>
<td>.56</td>
<td>.75</td>
<td>.17*</td>
</tr>
<tr>
<td>Organization</td>
<td>.74</td>
<td>.77</td>
<td>.34*</td>
</tr>
<tr>
<td>puts outline of lecture on blackboard</td>
<td>.59</td>
<td>.73</td>
<td>.22*</td>
</tr>
<tr>
<td>signals transition to next topic</td>
<td>.72</td>
<td>.88</td>
<td>.17*</td>
</tr>
<tr>
<td>Task orientation</td>
<td>.74</td>
<td>.80</td>
<td>.23*</td>
</tr>
<tr>
<td>states teaching objectives</td>
<td>.58</td>
<td>.67</td>
<td>.39*</td>
</tr>
<tr>
<td>sticks to point in answering questions</td>
<td>.51</td>
<td>.80</td>
<td>.19*</td>
</tr>
<tr>
<td>Interest</td>
<td>.74</td>
<td>.80</td>
<td>.23*</td>
</tr>
<tr>
<td>describes relevant personal experience</td>
<td>.58</td>
<td>.67</td>
<td>.39*</td>
</tr>
<tr>
<td>plays out practical applications</td>
<td>.51</td>
<td>.80</td>
<td>.19*</td>
</tr>
<tr>
<td>Rapport</td>
<td>.54</td>
<td>.83</td>
<td>.39*</td>
</tr>
<tr>
<td>offers to help students with problems</td>
<td>.66</td>
<td>.81</td>
<td>.43*</td>
</tr>
<tr>
<td>announces availability for consultation</td>
<td>.52</td>
<td>.69</td>
<td>.54*</td>
</tr>
<tr>
<td>shows concern for student progress</td>
<td>.62</td>
<td>.92</td>
<td>-.36*</td>
</tr>
<tr>
<td>Mannerisms</td>
<td>.68</td>
<td>.78</td>
<td>-.17*</td>
</tr>
<tr>
<td>avoids eye contact with students</td>
<td>.68</td>
<td>.78</td>
<td>-.17*</td>
</tr>
<tr>
<td>says “um” or “ah”</td>
<td>.55</td>
<td>.89</td>
<td>-.19*</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>.64</td>
<td>.64</td>
<td>-.48*</td>
</tr>
<tr>
<td>voice fades in mid-sentence</td>
<td>.70</td>
<td>.70</td>
<td>-.44*</td>
</tr>
<tr>
<td>stutters, mumbles, or slurs words</td>
<td>.62</td>
<td>.72</td>
<td>-.22*</td>
</tr>
</tbody>
</table>

* Significant at .05 level

Table 3T3 shows the type of results found in these preliminary studies. Actually what is shown here are results for a subset of 27 different teaching behaviors combined across six different studies carried out over a period of several years, some unpublished and some published (e.g., Murray, 1983a, 1985), with a total combined sample size of 424 teachers. Three important findings may be noted in Table 3. First, outside observers showed reasonably high interrater reliability in their recording of low-inference classroom teaching behaviors. The average interrater reliability in Table 3 is .77, suggesting that different observers working independently arrived at similar estimates of the frequency of occurrence of a given teaching behavior for a given instructor. Second, observer ratings of classroom teaching behaviors showed a clear factor structure. Table 3 shows the individual teaching behaviors that loaded highest on each factor in a principal-components factor analysis of the combined data set for 424 teachers. For example, teaching behaviors such as “uses concrete examples” and “stresses most important points” tended to correlate or cluster together to define a factor interpreted as Clarity, whereas “speaks expressively” and “shows facial expressions” were part of a cluster identified as Expressiveness. Although exactly the same factors were not found in all studies, usually around 8 to 10 clearly defined factors were identified in a given study, with Clarity, Expressiveness, Interaction, Organization, and Interaction disclosed in all or nearly all studies. Third, there were many significant correlations between specific teaching behaviors and student ratings of overall teaching effectiveness. A total of 18 of the 27 teaching behaviors listed in Table 3 correlated significantly with overall effectiveness rating. These were distributed across several different factors, but correlations tended to be highest and most consistent across studies for teaching behaviors loading on the Clarity, Expressiveness, and Interaction factors. Using multiple regression analysis, it was possible to account for 50 to 70 percent of the variance in student ratings of teaching with a set of as few as 10 low-inference teaching behaviors loading on Clarity, Expressiveness, and Interaction factors as predictor variables.

In summary, early research on low-inference teaching behaviors showed that highly rated university teachers do in fact teach differently (i.e., exhibit different classroom teaching behaviors) than less highly rated teachers. Highly rated teachers are more likely to do certain specific things in the classroom and less likely to do other things. One implication of these results is that college and university teachers can improve their teaching substantially by acquiring low-
inference teaching behaviors known to contribute to overall effectiveness ratings. Another important, but frequently overlooked implication of these results is that they provide an alternative or supplementary source of evidence in support of the validity of student ratings of teaching. Most of the traditional evidence cited in support of student ratings validity comes from studies in which instructors rated as effective by students are shown to be effective in other, more substantive ways, such as in terms of mean student performance on a common final exam in a multiple-section course (e.g., Cohen, 1981). From a teaching behaviors point of view, there are really two separate validity questions: (1) Do student ratings accurately reflect low-inference teaching behaviors of the instructor? and (2) Are these teaching behaviors significantly related to student cognitive and affective outcomes? The fact that student ratings can be predicted with considerable accuracy from observation of classroom teaching behaviors suggests that student ratings are highly valid in the first sense, in that they are determined by how the teacher actually teaches rather than by extraneous factors such as "popularity" or "personal warmth." Evidence from multi-section predictive validity studies, as well as other data to be reviewed below, suggests that student ratings are also moderately valid in the second sense, in that teaching behaviors are significantly related to student achievement.

Are Teaching Behaviors Related to Outcomes Other Than Student Ratings?

An important question arising from preliminary research reviewed above is whether low-inference classroom teaching behaviors are related to measures of teaching effectiveness other than student instructional ratings. If low-inference behaviors are related only to student ratings and not to measures such as student learning of course content, it could be argued that low-inference behaviors contribute only to the "popularity" of teaching and not to the "substance" of teaching. Table 4-14 shows the results of a study by Murray (1983b) in which the classroom teaching behaviors of 36 instructors in a multiple-section introductory Psychology course were studied in relation to 6 different measures of teaching effectiveness, including mean student ratings of overall teacher effectiveness and overall course quality, mean amount of studying per week reported by students, mean student performance on a common final exam, frequency of student registration in senior psychology courses, and mean student estimate of amount learned in the course. Please note that in this table, correlations are shown between outcome measures and teaching behavior factor scores rather than individual teaching behaviors, where factor scores were obtained by averaging frequency-of-occurrence ratings for all individual teaching behaviors that loaded .35 or higher on a given factor. For example, the factor score for "Rapport" was obtained by averaging across individual behaviors such as "addresses individual students by name," "talks with students before or after class," and "offers to help students with problems." As shown in the table, a total of 26 out of 78 correlations between teaching behavior factors and outcome measures were statistically significant, and multiple regression analyses showed that teacher behaviors collectively accounted for a substantial amount of variance in each outcome measure (ranging from 48% to 85%). In other words, it appears that low-inference teaching behaviors do indeed contribute to student outcomes other than instructional ratings. Specifically, the extent to which the student enjoys the course, studies a lot or a little, does well or poorly on the final examination, and enrolls in further courses in the same subject area appears to be determined, at least in part, by specific low-inference teaching behaviors of the instructor. On the other hand, relationships between teaching behaviors and outcome.
measures were not totally simple and straightforward, in that teaching behavior factors correlating significantly with one outcome measure often did not correlate similarly with other outcome measures. For example, Rapport correlated significantly with mean student rating of the teacher but not with mean performance on the common final examination, whereas Task Orientation correlated significantly with exam performance but not with teacher ratings. Enthusiasm, on the other hand, correlated significantly with both teacher rating and exam performance. This suggests that in order to be successful on a wide range of teaching outcome measures, an instructor must have a large and flexible repertoire of teaching behaviors!

Other research confirming that low-inference teaching behaviors correlate with outcomes other than student instructional ratings includes that of Hines, Crouch, and Kennedy (1985), Solomon, Rosenberg, and Bezelek (1986), and Tom and Cushman (1975). Hines, Crouch, and Kennedy investigated low-inference teaching behaviors in a sample of 32 student teachers enrolled in an experimental peer teaching program at Ohio State University. Each student was provided with a standard set of objectives and materials on the topic of matrix multiplication, and was allowed 2 days in which to prepare a 25-minute lesson to be presented to a group of 4 to 6 fellow students. Two observers viewed videotapes of the 32 mini-lessons and either counted or rated the frequency of occurrence of 29 low-inference teacher clarity behaviors. Additional data on teaching behaviors were obtained from student ratings and instructor self-ratings. Following instruction, students in each peer group rated their degree of satisfaction with the lesson and wrote a completion-type achievement test on matrix multiplication. Multiple regression analyses showed that observer estimates of low-inference clarity behaviors correlated strongly with both student instructional ratings and student performance on the achievement test, accounting for 36 percent of the variance in ratings and 52 percent of the variance in student achievement. The teacher clarity behaviors showing the strongest relationships to student ratings and student achievement included: “uses relevant examples,” “asks questions of students,” “reviews material,” “repeats points when students do not understand,” “teaches in a step-by-step manner,” and “provides frequent examples.” These behaviors are similar to items loading on the Clarity factor in the Murray (1983b) observational study reviewed above.

In summary, research evidence suggests that low-inference teaching behaviors do correlate with outcome measures other than student ratings, but correlations are not always consistent across different measures of teacher effectiveness.

---

**Is there a True Cause-Effect Relationship between Teaching Behaviors and Student Outcomes?**

Given that low-inference teaching behaviors appear to be correlated with several different measures of teaching effectiveness, including student learning, it is fair to ask whether these correlations reflect a true cause-effect relationship between teaching behaviors and student outcomes. A significant correlation between teaching behaviors and student performance may or may not reflect an underlying “forward causation” model in which teaching behaviors represent the “cause” (or one of the causes) and student learning the “effect.” Alternatively, causal models that could give rise to the same observed correlation include “backward causation,” in which teaching behaviors are the result of prior student learning rather than the cause of present learning; and “third-variable causation,” in which teaching behaviors covary with student learning because both are the result of some other variable. Establishing whether or not a forward cause-effect relationship exists between teaching behaviors and student outcomes is important not only for theoretical reasons but also for applied or practical reasons. From an applied point of view, it is counterproductive to encourage instructors to acquire or adopt particular teaching behaviors unless these behaviors have been shown to be causal antecedents of student learning. Adopting a teaching behavior that is a “correlate” but not a “cause” of student learning would fail to produce improved teaching effectiveness if the underlying causal pattern is either backward causation or third-variable causation.

Many procedures are available for clarifying the causal status of a correlation between teaching behaviors and student learning, but probably the best available option is to conduct a “true experiment,” under either laboratory or field conditions, in which teaching behaviors are experimentally manipulated while all other variables that could potentially affect student learning are held constant or controlled. If teaching behaviors continue to be significantly related to student learning under controlled experimental conditions, then the relationship can be assumed to reflect forward causation rather than backward or third-variable causation.

Evidence that there is indeed a true cause-effect relationship between teaching behaviors and student learning comes from research involving experimental manipulation of low-inference teaching behaviors. Figure 1 shows the results of a study by Murray (1978) in which randomly assigned groups of subjects viewed four different
organizational teaching behaviors, as did student performance on both multiple-choice and essay tests of amount learned. Given that extraneous variables were controlled in this study, the results suggest a true cause-effect relationship is operating between teaching behaviors and student outcomes. It is interesting to note that the positive impact of organizational teaching behaviors was more pronounced for the essay achievement test than for the multiple-choice test. This may suggest that organizational behaviors are particularly effective in helping students see the overall structure of topics and subtopics, and thus in transferring or applying knowledge to new situations.

Confirming evidence that low-inference teaching behaviors are causally related to student outcomes has been found in “true experiments” reported by Coats and Smidchens (1966), Land (1979), and Ware and Williams (1975), among others. Coats and Smidchens (1966) conducted a field experiment on the effects of teacher enthusiasm on student recall of lecture material. The subjects were 184 students enrolled in 8 introductory speech classes at the University of Michigan. Two instructors, graduate students in education, gave 10-minute guest lectures in 4 classes each. Lecture content was identical for all classes taught by a given instructor, but the lecture was presented in a “dynamic” fashion in two randomly selected classes, and in a “static” fashion in the two remaining classes. The dynamic condition included the usual behavioral ingredients of teacher enthusiasm: movement, gestures, eye contact with students, vocal inflection, and minimal reliance on lecture notes. The static lecture was presented with good diction and volume, but was read verbatim from a manuscript, and included a minimum of eye contact, vocal inflection, and animation. Immediately following the lecture, students in all classes completed a 10-item multiple-choice test based on lecture material. It was found that students in the dynamic condition performed better on the recall test than students in the static condition. Mean recall was approximately 20 percent higher for dynamic lectures than for static lectures, and teacher enthusiasm accounted for 36 percent of the total variance in student recall scores.

Land (1979) evaluated the combined impact of 6 low-inference teacher clarity variables, including vagueness terms (i.e., inexact statements), transition signals (i.e., cues that mark the end of one topic and the beginning of another), and verbal mazes (i.e., false starts or halts in speech). Two versions of the same videotaped lecture, varying only in the presence or absence of the 6 clarity behaviors, were presented to students in an introductory education course. Students were told
to take notes during the lesson, but were not allowed to use their notes during subsequent testing. The achievement test, consisting of 30 multiple-choice items written at the comprehension level of Bloom's taxonomy, was completed either immediately following the lesson or after a delay of one week. Statistical analysis showed that, for both immediate and delayed testing, comprehension scores were higher for the high-clarity lesson than for the low-clarity lesson. Land hypothesized that the use of explicit transition signals in classroom teaching assists students in organizing the subject matter, whereas the use of vague terms causes students to lose confidence in the instructor and thus in themselves.

In summary, evidence from both field and laboratory experiments suggests that the relationship between low-inference teaching behaviors and student learning generally conforms to a "forward causation" model in which teaching behaviors are the "cause" and student learning is the "effect." It should be noted, however, that some experiments (e.g., Anderson and Withrow, 1981) have failed to find any sort of significant cause-effect relationship between teaching behaviors and student learning.

**WHAT ARE THE PROCESSES OR MECHANISMS UNDERLYING THE RELATIONSHIP BETWEEN TEACHING BEHAVIORS AND STUDENT LEARNING?**

Given that teaching behaviors appear to be related to a wide range of student outcomes, including student learning, and given that this relationship appears to reflect a forward cause-effect pattern, what are the psychological processes or mechanisms that mediate this relationship? In other words, what are the cognitive or affective processes occurring in the student that give rise to the positive impact of teaching behavior X upon student learning? Among other things, it is possible that teaching behavior X causes students to perceive differences more accurately, to form more structured knowledge representations, to experience less anxiety, or to develop an improved self-concept. The question of underlying processes is of course a theoretical question, but unfortunately research on low-inference teaching behaviors has tended to be rather atheoretical in nature. Most research to date has attempted to demonstrate empirical relationships between teaching behaviors and student outcomes rather than to identify processes or mechanisms underlying these relationships.

**Although we still know very little about the psychological processes intervening between teaching behaviors and student learning, some promising leads in this area have been provided by Ray Perry's research on student attributions and perceived control in relation to classroom teaching behaviors, and by recent doctoral dissertations completed by Dieter Schonwetter at the University of Manitoba and Andrea Wood at the University of Western Ontario.**

Research by Perry and colleagues, summarized in Perry (1991) has demonstrated that teacher expressiveness, defined by low-inference behaviors such as body movement, vocal variation, eye contact, and humor, has a stronger impact on student learning for students with internal success and failure attributions who perceive themselves to be in control of their environments, than for students with external attributions or lack of perceived control. This result suggests that lack of perceived control may interfere with processes normally activated by teacher expressiveness, such as increased attention to lecture material; or alternatively, that teacher expressiveness may facilitate student learning in part by creating a stronger sense of perceived control in students.

Schonwetter's (1996) doctoral research included a laboratory experiment assessing the interaction of two different categories of teaching behavior in determining student attention and learning: (1) teacher expressiveness, defined by eye contact, body movement, hand gestures, and use of humor, and (2) teacher organization, defined by providing an outline, using headings, and signalling topic transitions. A total of 380 students viewed one of four different versions of a videotaped economics lecture involving all possible combinations of low vs. high teacher expressiveness with low vs. high teacher organization (2 x 2 design). Student selective attention to lecture material was measured by self-report ratings and by a free recall test, whereas student learning was measured by post-lecture multiple-choice achievement tests assessing both recognition and application of concepts, and by self-ratings of amount learned. It was found that teacher organization had strong and significant effects on both student attention and student learning, whereas teacher expressiveness had weak and generally nonsignificant effects on both attention and learning. This result suggests (1) that low-inference teacher organization behaviors may be a more basic or prepotent factor in teaching effectiveness than low-inference teacher expressiveness behaviors, (2) that teacher organization may influence student learning by way of selective attention, and (3) that teacher expressiveness may affect student learning only if
the teacher is organized, rather than vice versa as suggested by Murray (1983a). Schouwester points out, however, that the differentiation of low and high teacher expressiveness conditions in this study may not have been clear enough to result in statistically significant effects on student attention or learning.

Andrea Wood's (1998) doctoral thesis attempted to decide among three different models of the underlying process by which teacher enthusiasm affects student learning, namely selective attention, motivation to learn, and improved memory encoding. The selective attention model suggests that enthusiastic teaching behaviors such as body movement and vocal expressiveness improve learning by eliciting student attention to lecture material as opposed to distracting stimuli. The motivation model states that, perhaps through imitation or modeling on the part of student, enthusiastic teaching behaviors cause students to develop higher levels of motivation to learn the subject matter, both inside and outside the classroom. The memory encoding model assumes that expressive teaching behaviors improve learning by signaling important ideas in spoken text, and thus help students understand the overall structure or “meaning” of the subject matter.

To test among these three models, Wood conducted a laboratory experiment with videotaped lectures in a simulated classroom, where teacher enthusiasm was defined as the occurrence of the following low-inference teaching behaviors: vocal variation, movement and gesture, pausing to emphasize points, humor, facial expression, and eye contact. Three hundred introductory psychology students were randomly assigned to treatment groups receiving four different versions of a 16-minute lecture on memory theory, namely Low Enthusiasm, High Enthusiasm/Strategic, High Enthusiasm/Random, and High Enthusiasm/Uniform. The Low Enthusiasm lecture included few if any enthusiastic teaching behaviors. The High Enthusiasm/Strategic lecture included a high frequency of enthusiastic teaching behaviors, and these low-inference behaviors were properly coordinated with the topic structure of the lesson. The High Enthusiasm/Random condition also included frequent use of enthusiastic behaviors, but these behaviors sometimes did and sometimes did not coincide with the topic structure of the lesson. The High Enthusiasm/Uniform condition featured frequent use of enthusiastic teaching behaviors, but their occurrence remained constant at all points throughout the lecture.

During the experiment, each of the three hypothesized mediating processes, namely selective attention, motivation, and memory encoding, was monitored by at least two different indicator variables. Selective attention was measured by: (1) reaction time to a secondary tone detection task presented at random intervals during the videotaped lecture, and (2) observed frequency of on-task attending behavior during the lecture. It was expected that teacher enthusiasm would lead to slower secondary task reaction times, indicating greater attention to the lecture, and more frequent on-task behaviors. Student motivation was measured by: (1) a questionnaire assessing degree of interest in the lecture topic, and (2) observed frequency of written requests from students for mailed reading material relevant to the lecture. Memory encoding was measured by: (1) overall recall, defined as the total number of lecture propositions correctly recalled, (2) topic access, defined as the number of lecture topics for which at least one proposition was recalled, (3) conditional recall, the percentage of propositions recalled per topic, given that at least one proposition was recalled, and (4) topic representation, defined as the degree of similarity (rank order correlation) between the subject's order of recalled topics and the actual order of presentation of topics in the lecture. The last three memory measures were intended to assess the extent to which participants had encoded the topical structure of the lecture as a whole, and as outlined further below, were expected to be facilitated by strategic, but not by random or uniform use of teacher enthusiasm.

After viewing the videotaped lecture, subjects completed a multiple-choice test assessing learning of lecture content, and provided ratings of quality of teaching.

Table 3T5 summarizes the main results of the experiment, namely mean scores for the four treatment conditions on each of eight variables assessing potential mediators (2 attention, 2 motivation, 4 memory encoding) and each of two outcome measures (student learning, student ratings of lecture quality). It may be noted that subjects in the three High Enthusiasm conditions (Strategic, Random, and Uniform) tended to score higher than subjects in the Low Enthusiasm condition on all indicator variables assessing potential mediators, as well as on learning of lecture content and ratings of lecture quality. For example, subjects in the three High Enthusiasm conditions showed slower reaction times to the secondary task (indicating greater attention to the lecture), higher scores on the questionnaire assessing motivation for further learning, and better encoding of the topic structure of the lecture (as reflected in topic representation scores). These results suggest that any or all of the three hypothesized mediating processes (i.e., selective attention, motivation to learn, or memory encoding) could be responsible for the positive impact of teacher enthusiasm on
### Table 5: Group Means Scores on Indicator Variables and Outcome Measures (Wood, 1990)

<table>
<thead>
<tr>
<th>Treatment Conditions</th>
<th>High Enthus Strategic</th>
<th>High Enthus Random</th>
<th>High Enthus Uniform</th>
<th>Low Enthus</th>
<th>Univariate F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selective Attention</td>
<td>817.28</td>
<td>829.85</td>
<td>895.25</td>
<td>471.26</td>
<td>107.14*</td>
</tr>
<tr>
<td>Reaction Time</td>
<td>98.25</td>
<td>98.24</td>
<td>95.69</td>
<td>69.16</td>
<td>125.44*</td>
</tr>
<tr>
<td>On-Task Behavior</td>
<td>2.52</td>
<td>2.23</td>
<td>1.91</td>
<td>1.83</td>
<td>9.72*</td>
</tr>
<tr>
<td>Motivation</td>
<td>.24</td>
<td>.25</td>
<td>.17</td>
<td>.09</td>
<td>6.37</td>
</tr>
<tr>
<td>Request for Further</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory Encoding</td>
<td>6.92</td>
<td>5.21</td>
<td>3.11</td>
<td>5.04</td>
<td>17.52*</td>
</tr>
<tr>
<td>Overall Recall</td>
<td>3.64</td>
<td>2.87</td>
<td>1.99</td>
<td>2.87</td>
<td>13.10*</td>
</tr>
<tr>
<td>Topic Access</td>
<td>23.80</td>
<td>22.59</td>
<td>15.63</td>
<td>20.82</td>
<td>10.32*</td>
</tr>
<tr>
<td>Conditional Recall</td>
<td>72.72</td>
<td>71.51</td>
<td>58.58</td>
<td>35.35</td>
<td>5.08*</td>
</tr>
<tr>
<td>Topic</td>
<td>8.76</td>
<td>6.79</td>
<td>4.48</td>
<td>6.05</td>
<td>21.05*</td>
</tr>
<tr>
<td>Representation</td>
<td>5.21</td>
<td>4.64</td>
<td>4.78</td>
<td>2.51</td>
<td>55.00*</td>
</tr>
</tbody>
</table>

* Significant at .01 level

Student memory and learning scores should be facilitated equally by all three High Enthusiasm conditions. Converging evidence favoring the memory encoding model over the attention and motivation models of teacher enthusiasm was obtained from multiple regression analyses testing the role of selective attention, motivation, and memory encoding variables in mediating the relationship between teacher enthusiasm (dummy coded) and student learning. It was found that the correlation between teacher enthusiasm and student learning was changed significantly only when memory encoding was entered or removed from the regression equation, indicating that although all three mediators may be involved to some degree, memory encoding plays a more decisive role in mediating the relationship between teacher enthusiasm and student learning than either selective attention or motivation. Although the memory encoding model generally provides a better account of the Wood findings than either the attention or the motivation model, it is not without problems. For example, it fails to account for the fact that conditional recall scores did not differ in Strategic vs. Random vs. Low Enthusiasm conditions, or for the fact that ratings of lecture quality were equally high in all three High Enthusiasm conditions.

The results of the Wood study provide preliminary evidence as to the specific underlying processes or mechanisms that mediate the relationship between teacher enthusiasm and student learning. Although all three hypothesized mediating processes (selective attention, motivation to learn, and memory encoding) may be involved to some degree in mediating the relationship between teacher enthusiasm and student learning, it appears that memory encoding plays a more decisive role than selective attention or motivation. Figure 2 shows one possible way in which memory encoding may interact with selective attention and motivation to mediate the effect of teacher enthusiasm on student learning. According to this model, teacher enthusiasm will facilitate student learning only if it activates memory encoding of text structure either directly or indirectly by way of selective attention. Reflecting its limited significance in the present study, motivation is treated as a secondary outcome or by-product of student learning in this model. One interesting implication of the Wood results is that teacher enthusiasm does not inevitably facilitate student learning in all cases. In order to contribute positively to student learning, enthusiastic teaching behaviors must be used strategically so as to emphasize the topic structure of the material to be learned.

In summary, recent research has provided preliminary evidence regarding the processes or mechanisms underlying the relationship...
between low-inference teaching behaviors and student learning. However, only a few studies have been conducted to date, and these have tended to focus on only one type of teaching behavior, namely expressiveness or enthusiasm. One notable consistency between the Schonwetter (1996) and Wood (1998) studies is that in both cases, organizational teaching behaviors were found to be more basic or fundamental than expressive teaching behaviors, in the sense that teacher expressiveness was effective only to the extent that it was preceded by or operated through structuring of content. Further support for this hypothesis comes from a meta-analyses reported by Feldman (1989, 1997) in which teacher organization, on average, correlated higher with student performance on common final exams in multiple-section courses than did teacher expressiveness.

**IS THE RELATIONSHIP BETWEEN LOW-INFERENCE BEHAVIORS AND STUDENT INSTRUCTIONAL RATINGS CONSISTENT ACROSS DIFFERENT SITUATIONS OR CONTEXTS?**

I now want to turn to some recent studies investigating the controversial question of whether or not the impact of low-inference teaching behaviors is consistent across different contexts or situations. Some writers do not consider this question to be controversial. They assert, sometimes with and sometimes without supporting empirical evidence, that teaching is contextual or context-dependent. Thus, teaching behaviors have no generality across situations, or what works in one situation will not necessarily work in another situation. For example, Brookfield (1990, page 12) reported that, "...every context in which I worked contained factors that prevented the neat application of principles and techniques of ‘good’ practice." Similarly, Shulman (1986) criticized teacher effectiveness research for failing to consider differences in subject matter, and for assuming "generic" teaching behaviors that are applicable in all contexts. Shehan (1975) argued against the use of standardized teacher evaluation forms on the grounds that the characteristics assessed by such forms are not equally relevant in all contexts or for all types of students. Finally, Good and Brophy (1990, page 286) concluded from a review of teacher effectiveness research that, "Few if any instructional behaviors are appropriate in all teaching contexts,..."

Contrary to this popular line of thought, some researchers have reported that the impact of teacher characteristics is surprisingly consistent across different contexts or situations. Marques, Lane, and Dorfman (1979) asked students and faculty in four different academic divisions (humanities, engineering, natural sciences, social sciences) to rate the overall effectiveness of hypothetical instructors who varied systematically on specified dimensions. It was found that there was strong consensus between students and faculty and among respondents from different academic fields as to the relative importance of each instructional component or dimension in determining overall effectiveness, suggesting that perceived teaching effectiveness tends to be "transitionally invariant." Similarly, Pohlmann (1976) found that the correlation of overall teacher effectiveness ratings with specific instructional practices (e.g., encouraging student participation, specifying course objectives) did not differ significantly across five academic disciplines, indicating that what makes an effective teacher is basically the same regardless of what is taught. Finally, Roberts and Becker (1976) conducted an observational study of 123 instructors involved in one-to-one teaching of industrial and technical skills in high schools and community colleges, and found that the teaching behaviors differentiating between effective and ineffective instructors were remarkably similar to those reported for traditional large-class teaching.

Most of the previous research on the situational dependency of teaching has focused on high-inference rather than low-inference teacher characteristics, has relied on student raters to assess both teacher characteristics and overall teacher effectiveness (thus leading to possible spurious correlations due to judgement bias), and has dealt mainly with only one type of context variable, namely academic field. The studies reviewed below focus on specific low-inference teaching behaviors, use trained classroom observers to assess teaching behaviors, and report data on three different contextual variables,
namely academic discipline, class size, and teacher gender. Also, these studies investigate the contextual dependency of low-inference teaching behaviors with respect to two different measures: (1) frequency of occurrence—e.g., does Behavior X occur more frequently in small classes than in large classes? and (2) correlation with outcome measures—e.g., does Behavior Y correlate higher with student ratings for Chemistry teachers than for Sociology teachers?

ACADEMIC DISCIPLINE

The first study, which compared teaching behaviors across three different academic discipline areas at the University of Western Ontario, was done in collaboration with Robert Renaud (Murray and Renaud, 1995). An aggregated sample of 401 teachers was obtained by combining data from seven previous studies involving classroom observation of faculty members teaching undergraduate lecture—or lecture discussion courses at the University of Western Ontario. In all of these studies, classroom observers summarized their 3 hours of classroom observation of a given teacher on the same 100-item version of the Teacher Behaviors Inventory. Overall teaching effectiveness was measured by end-of-term student ratings on a standardized teaching evaluation form administered in the same course and semester in which classroom observation took place. The three disciplinary groupings compared in this study were: Arts and Humanities (N = 117), Social Sciences (N = 149), and Science and Mathematics (N = 135). Mean observer ratings were calculated for each teacher on each of the 100 TBI items, then teaching behavior factor scores were obtained for each teacher by averaging TBI ratings across the teaching behaviors loading highest on each of 10 factors.

Table 6 shows the mean rated frequency of occurrence of teaching behaviors, or more accurately, teaching behavior factor scores, for Arts and Humanities versus Natural Science versus Social Science teachers. Statistical analysis of these data showed that 6 of 10 categories of teaching behavior differed significantly in frequency of occurrence across the three disciplinary groups, namely Interaction, Organization, Pacing, Disclosure, Rapport, and Mannerisms. For example, Arts and Humanities teachers were more likely than Social Science and Natural Science teachers to show behaviors in the Rapport and Interaction categories (e.g., encouraging student participation, addressing individual students by name), whereas teachers in Social Science and Natural Science were more likely than Arts and Humanities teachers to show behaviors loading on the Organization and Pacing dimensions (e.g., putting outline on blackboard, sticking to the point in answering questions).

Table 7 compares Arts vs. Natural Science vs. Social Science teachers in terms of the magnitude and direction of correlations between teaching behavior factors and end-of-term student ratings of overall teaching effectiveness. As in previous research, all 10 teaching behavior dimensions showed generally positive correlations with student ratings of overall teaching effectiveness. Furthermore, and more relevant to the present issue, correlations between teaching behaviors and student ratings tended to be similar across academic fields. Statistical analysis of these data (using Fisher's r to z transformation and Fisher's test of the significance of differences between correlation coefficients, see McNemar, 1962, pages 139–140) indicated that of the 30 possible pairwise differences between academic disciplines in Table 7 (3 possible pairwise comparisons for each of 10 teaching behavior factors), only two were statistically significant. Specifically, Rapport correlated higher with overall teaching effectiveness ratings for both Social Science and Natural Science teachers than for Arts and Humanities teachers.

In summary, the results of this study suggest that teachers of different academic disciplines at the University of Western Ontario differed in the frequency with which they exhibited various low-inference teaching behaviors, but did not differ in the correlation of
Table 7: Correlation of Teaching Behaviors With Overall Teacher Effectiveness Ratings in Different Disciplinary Groups (Murray & Renand, 1995)

<table>
<thead>
<tr>
<th>Teaching Behavior Factor</th>
<th>Disciplinary Group</th>
<th>Arts and Humanities</th>
<th>Social Sciences</th>
<th>Natural Sciences and Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>.498</td>
<td>.562</td>
<td>.647</td>
<td></td>
</tr>
<tr>
<td>Expressiveness</td>
<td>.308</td>
<td>.402</td>
<td>.446</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>.417</td>
<td>.441</td>
<td>.502</td>
<td></td>
</tr>
<tr>
<td>Organization</td>
<td>.359</td>
<td>.361</td>
<td>.439</td>
<td></td>
</tr>
<tr>
<td>Pacing</td>
<td>.511</td>
<td>.496</td>
<td>.609</td>
<td></td>
</tr>
<tr>
<td>Disclosure</td>
<td>.254</td>
<td>.305</td>
<td>.221</td>
<td></td>
</tr>
<tr>
<td>Interest</td>
<td>.352</td>
<td>.357</td>
<td>.433</td>
<td></td>
</tr>
<tr>
<td>Rapport*</td>
<td>.316</td>
<td>.391</td>
<td>.579</td>
<td></td>
</tr>
<tr>
<td>Manerisms</td>
<td>.513</td>
<td>.455</td>
<td>.258</td>
<td></td>
</tr>
<tr>
<td>Speech Quality</td>
<td>.406</td>
<td>.615</td>
<td>.650</td>
<td></td>
</tr>
</tbody>
</table>

*One or more pairwise differences between disciplinary groups are significant at .05 level.

These teaching behaviors with student ratings of teaching effectiveness. For example, although teaching behaviors loading on the Organization and Pacing factors occurred more frequently in Science and Social Science teachers than in Arts teachers, the extent to which Organization and Pacing behaviors (and most other teaching behaviors) "paid off" or contributed to overall effectiveness ratings was essentially the same for Social Science teachers as for Arts teachers. This result needs to be replicated in other teacher samples at other institutions. If so replicated, this result would suggest that the teaching behaviors that contribute to successful teaching are surprisingly similar in different academic disciplines, and would run directly counter to the widely shared belief that teaching effectiveness is highly context-dependent.

One practical implication of the above findings is that, contrary to the claim that we need to design separate and distinctively different faculty development and faculty evaluation programs for different academic fields (e.g., Sheehan, 1975), it may not be unreasonable, given that the correlation of teaching behaviors with overall effectiveness may be similar across academic disciplines, to offer the same types of teaching improvement programs for faculty in all disciplines, or to use a common form for student evaluation of teaching in all faculties or

The same aggregated data set used above in comparing disciplinary groups provided data for a second study comparing samples of female (N = 60) and male (N = 364) teachers in terms of both frequency of occurrence of low-inference teaching behaviors and correlation of teaching behaviors with overall effectiveness ratings. It may be noted that females constituted about 14 percent of the aggregated sample, which is roughly equal to the percentage of female faculty at the University of Western Ontario at the time that this study was conducted. Results are summarized in Tables 8 and 9. Statistical analysis of the data in Table 8 indicated that 4 of 10 categories of teaching behavior showed significant differences in frequency of occurrence for female vs. male teachers, namely Interaction, Pacing, Disclosure, and Rapport. In general, it appears that female teachers were more likely to show behaviors such as encouraging student participation and showing concern for student progress, whereas male teachers were more likely to cover material at a rapid pace and provide information about tests and assignments. However, analysis of the data in Table 9 showed that the correlation of teaching behavior factors with teacher effectiveness rating differed significantly in female vs. male teachers for only one of 10 teaching behavior factors. As

Table 8: Mean Rated Frequency of Occurrence of Teaching Behaviors as a Function of Teacher Gender

<table>
<thead>
<tr>
<th>Teaching Behavior Factor</th>
<th>Teacher Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female</td>
</tr>
<tr>
<td>Clarity</td>
<td>3.54</td>
</tr>
<tr>
<td>Expressiveness</td>
<td>3.27</td>
</tr>
<tr>
<td>Interaction*</td>
<td>3.33</td>
</tr>
<tr>
<td>Organization</td>
<td>3.09</td>
</tr>
<tr>
<td>Pacing*</td>
<td>3.83</td>
</tr>
<tr>
<td>Disclosure*</td>
<td>3.15</td>
</tr>
<tr>
<td>Interest</td>
<td>3.08</td>
</tr>
<tr>
<td>Rapport*</td>
<td>3.91</td>
</tr>
<tr>
<td>Manerisms</td>
<td>4.13</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>4.09</td>
</tr>
</tbody>
</table>

* Difference between gender groups is significant at .05 level.
Table 9: Correlation of Teaching Behaviors With Overall Teacher Effectiveness Rating as a Function of Teacher Gender

<table>
<thead>
<tr>
<th>Teaching Behavior Factor</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>.530</td>
<td>.575</td>
</tr>
<tr>
<td>Expressiveness</td>
<td>.531</td>
<td>.377</td>
</tr>
<tr>
<td>Interaction</td>
<td>.521</td>
<td>.407</td>
</tr>
<tr>
<td>Organization</td>
<td>.445</td>
<td>.373</td>
</tr>
<tr>
<td>Pacing*</td>
<td>.267</td>
<td>.531</td>
</tr>
<tr>
<td>Disclosure</td>
<td>.397</td>
<td>.336</td>
</tr>
<tr>
<td>Interest</td>
<td>.601</td>
<td>.464</td>
</tr>
<tr>
<td>Rapport</td>
<td>.623</td>
<td>.504</td>
</tr>
<tr>
<td>Mannerisms</td>
<td>.314</td>
<td>.440</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>.548</td>
<td>.590</td>
</tr>
</tbody>
</table>

* Difference between gender groups is significant at .05 level.

may be noted in the table, Pacing correlated significantly higher with overall effectiveness for males than for females. In summary, although male and female teachers differed in the frequency with which they exhibited certain teaching behaviors, they generally did not differ in the extent to which these behaviors correlated with or contributed to student ratings of overall teaching effectiveness. Consistent with what was found for academic disciplines, we again have evidence that teaching behaviors differed across situations or contexts (i.e., genders) in frequency of occurrence, but did not differ across contexts in correlation with student ratings of teaching. The low-inference classroom teaching behaviors that contributed to good teaching for male teachers are more or less the same as the teaching behaviors that contributed to teaching effectiveness for female teachers. Contrary to Basow and Distentefeld’s (1985) laboratory-based finding that expressive teaching behaviors contributed more positively to student evaluation of teaching for male teachers than for female teachers, the present data indicate no significant gender difference in the correlation between teacher expressiveness and student ratings under field conditions (and in fact, the direction of the difference, although nonsignificant, is actually opposite to that reported by Basow and Distentefeld). As with the previously reported data for academic disciplines, the present results need to be replicated with other samples of teachers in other institutions.

CLASS SIZE

Research testing the consistency of teaching behaviors in small vs. large classes was reported in a Masters thesis by Andrea Wood (1994) at the University of Western Ontario. She studied a sample of 38 psychology professors, 19 of whom were teaching a large lecture-type class (mean size = 217, range = 25 to 400), and 19 of whom were teaching a small, seminar style class (mean size 21, range 7 to 40) during the same academic year. Similar to the academic discipline and gender data reported previously, this study found differences between small and large classes in the frequency of occurrence of low-inference teaching behaviors, but not in the correlation of teaching behaviors with student ratings of overall teaching effectiveness. Tables 1011 and 1111 summarize the frequency and correlation results respectively. It may be noted that the teacher behavior factor structure in this study was not exactly the same as in previous studies, in that only 8 rather than 10 factors were defined. Statistical analysis of the data in Table 10 indicated significant differences between small and large classes for 3 of the 8 teaching behavior factors. Specifically, Interaction and Rapport behaviors more frequent for teachers of small seminar classes than for teachers of large lecture classes, whereas Organization behaviors were more frequent in large classes than in small classes. These results are not terribly surprising, and are more or less what Wood had predicted. However, what came as a big surprise was the fact that of the 8 teaching behavior factors listed in Table 11, none differed significantly across

Table 10: Mean Rated Frequency of Occurrence of Teaching Behaviors as a Function of Class Size (Wood, 1994)

<table>
<thead>
<tr>
<th>Teaching Behavior Factor</th>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>3.59</td>
<td>3.62</td>
</tr>
<tr>
<td>Enthusiasm</td>
<td>3.39</td>
<td>3.47</td>
</tr>
<tr>
<td>Interaction*</td>
<td>3.62</td>
<td>3.30</td>
</tr>
<tr>
<td>Organization</td>
<td>3.24</td>
<td>3.62</td>
</tr>
<tr>
<td>Pacing</td>
<td>3.83</td>
<td>3.79</td>
</tr>
<tr>
<td>Disclosure</td>
<td>3.37</td>
<td>3.46</td>
</tr>
<tr>
<td>Rapport*</td>
<td>4.00</td>
<td>3.43</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>4.17</td>
<td>4.00</td>
</tr>
</tbody>
</table>

* Difference between small and large groups significant at .05 level.
Table 1: Correlation of Teaching Behaviors with Overall Effectiveness Ratings as a Function of Class Size (Wood, 1994)

<table>
<thead>
<tr>
<th>Teaching Behavior Factor</th>
<th>Small</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarity</td>
<td>.75</td>
<td>.83</td>
</tr>
<tr>
<td>Enthusiasm</td>
<td>.63</td>
<td>.41</td>
</tr>
<tr>
<td>Interaction</td>
<td>.23</td>
<td>.07</td>
</tr>
<tr>
<td>Organization</td>
<td>.57</td>
<td>.73</td>
</tr>
<tr>
<td>Pacing</td>
<td>.73</td>
<td>.73</td>
</tr>
<tr>
<td>Disclosure</td>
<td>.42</td>
<td>.55</td>
</tr>
<tr>
<td>Rapport</td>
<td>.07</td>
<td>.24</td>
</tr>
<tr>
<td>Speech Quality</td>
<td>.73</td>
<td>.59</td>
</tr>
</tbody>
</table>

* Difference between small and large groups significant at .05 level.

small seminar vs. large lecture classes in terms of correlation with overall effectiveness ratings. It was expected that interaction behaviors would not only occur more frequently in small than in large classes, but would contribute more to overall teaching effectiveness in small classes. Similarly, it was expected that organizational behaviors would both occur more often and correlate higher with overall effectiveness in lecture classes than in seminar classes. Contrary to expectation, neither interaction behaviors nor organization behaviors (nor any other type of teaching behavior) differed significantly between small and large classes in correlation with overall teacher ratings. So again we have suggestive evidence that although teaching behaviors may differ across situations or contexts in frequency of occurrence, they do not differ across situations in their correlation with student instructional ratings.

Although small class teachers may be more likely to exhibit interactive teaching behaviors than large class teachers, the extent to which interactive behaviors "pay off" in higher teacher ratings may be just as high for large classes as it is for small classes. Similarly, although large class teachers tend to show organizational teaching behaviors more frequently, these behaviors may contribute to perceived teaching effectiveness just as much for small classes as for large classes. One possible criticism of the Wood study is that, contrary to instructor self-reports, the method or style of instruction used in small and large classes may in fact have been very similar or identical (for example, it is possible that teachers used 100 percent lecturing in both types of classes). Thus, what appeared to be a contextual difference may not have been a difference at all. This argument has intuitive appeal, but it will not account for the fact that there were significant differences in the frequency of teaching behaviors across small vs. large classes in the Wood study.

In summary, we have evidence from three separate studies that low-inference teaching behaviors showed surprising consistency across situations or contexts in their correlation with overall teaching effectiveness ratings. This finding runs counter to the popular view that what constitutes good teaching is embedded in context and varies systematically from one context or situation (e.g., Brookfield, 1990; Good and Brophy, 1990; Shulman, 1986). The "contextual dependency" view implies that the correlation between frequency of occurrence of specific teaching behaviors and measures of overall teaching effectiveness should differ significantly in different contexts. An alternative view, favored by the present author, is that there may be certain fundamental or generic teaching behaviors such as expressiveness, organization, clarity of explanation, and encouragement of student participation, that contribute more or less equally to teaching effectiveness in all or nearly all contexts ranging from teaching reading in Grade 1 to teaching biochemistry to graduate students. It makes sense to me that there may be certain key low-inference teaching behaviors that convey enthusiasm for the subject matter or contribute to clarity of explanation and are equally effective in all contexts. On the other hand, it is possible that there may be other low-inference teaching behaviors that do in fact differ significantly across contexts in their impact on teaching effectiveness. Obviously, this is a question that requires further investigation, and one can only hope that the eventual answer to this question be based on systematic empirical evidence rather than totally on anecdote or personal opinion.

**Can Research on Low-Inference Teaching Behaviors Be Applied Successfully to Improvement of Teaching?**

I want to turn now to some studies that have attempted to apply research on low-inference teaching behaviors to improvement of teaching. Given that there appear to be clear cause-effect relationships between specific classroom teaching behaviors and measures of teaching effectiveness, it would seem that university teachers could improve their effectiveness by acquiring or emulating some of the teaching behaviors found to be important in research. But we all know that there are at least two things in this world that are much easier said than done: (1) applying research findings to real world problems,
and (2) acquiring new teaching behaviors. I will briefly review two studies that attempted to improve teaching by modifying low-inference teaching behaviors, one by way of feedback and one by way of training.

**Behavioral Feedback**

One way of applying research on low-inference teaching behaviors is by developing better procedures for providing formative or diagnostic feedback to instructors on their classroom teaching. One obvious way of providing such feedback is by way of student ratings, but the typical student rating form in current use focuses primarily on global, high-inference teacher characteristics, and therefore is less ideal for purposes of formative feedback. For example, a poor rating on a high-inference item such as “explains clearly” or “is well prepared” may signify that there is a problem with clarity or with preparation, but provides no hint as to specifically what the problem is, what is causing it, and what needs to be done to bring about improvement. Over the years I have experimented with various “formative” or “diagnostic” student rating forms that focus on specific, low-inference teaching behaviors such as “maintains eye contact with students” and “signals the transition from one topic to the next,” and thus are intended to provide the instructor a much clearer signal as to what is wrong and what remedial action is needed.

Despite the intuitive appeal of this goal, early attempts to demonstrate beneficial effects of low-inference behavioral feedback from students (e.g., McLear, 1979; Froman and Owen, 1980; Creighton, 1990) met with very little success. One possible reason was the problem of finding a way to present behavioral feedback to faculty members in a succinct, understandable format. One thing we discovered in these early studies was that a computer printout showing means, standard deviations, and percentile scores for 100 teaching behaviors causes the eyes of most faculty members to glaze over instantly!

Murray and Smith (1989) constructed and evaluated a new, revised behavioral feedback form that was intended to solve these problems. Essentially this is a formative feedback version of the previously described Teacher Behavior Inventory. The feedback version of the TBI consists of 50 or 60 items, each referring to a specific classroom teaching behavior that has been found in previous research to show adequate interrater reliability and to correlate significantly with student evaluation of overall teacher effectiveness. The difference is that the feedback version of the TBI is completed by students rather than outside observers and uses a very different type of rating scale. Instead of rating the frequency of occurrence of teaching behaviors, students rate each teaching behavior on a bipolar 5-point scale to indicate whether, for purposes of improving teaching, the behavior in question needs to be increased in frequency of occurrence (rating of +1 or +2), decreased in frequency of occurrence (rating of −1 or −2), or unchanged in frequency of occurrence (zero rating). The feedback version of the TBI is intended to provide behavioral feedback that is simple, direct, easy to interpret, and obvious in its implications for improvement. Instructors can obtain feedback on which teaching behaviors to change, and in what direction, simply by identifying TBI items where student ratings deviate noticeably from zero and have relatively small standard deviations.

To determine whether the revised TBI can in fact contribute to improvement of teaching, Murray and Smith (1989) conducted an experiment in which the instructors were 60 graduate students serving as teaching assistants in the Departments of Geography, Psychology, and English at the University of Western Ontario. Half of the instructors in each discipline were randomly assigned to receive midterm TBI feedback from students in their courses (experimental group), whereas half were assessed with the TBI at midterm but did not receive feedback of the results of this assessment (control group). Behavioral feedback consisted of the mean and standard deviation of student ratings for each item of the diagnostic TBI, plus brief instructions for interpreting the data provided. The experiment was conducted over an 8-month (September to April) academic term, with behavioral feedback provided at the approximate midpoint (late December). Thus the post-feedback interval was approximately 4 months, which is considerably longer than that used in most previous studies. Improvement of classroom teaching was measured by amount of pretest-to-posttest gain in student ratings of overall teaching effectiveness.

As may be noted in Figure 3, midterm behavioral feedback led to significant improvement of classroom teaching in all three academic departments, as indicated by significantly larger pretest (mid-term) to posttest (end-of-term) gains in student ratings of overall teaching effectiveness for experimental group instructors receiving behavioral feedback than for control instructors. Furthermore, the estimated effect size for behavioral feedback across all three departments was approximately .73 standard deviation units, which is considerably higher than the average effect size of .20 reported by Cohen (1980) for feedback from traditional high-inference student evaluation forms.
So, despite the pessimistic outcomes of earlier research, the Murray and Smith (1989) study suggests that, under the right conditions, feedback on low-inference teaching behaviors can contribute significantly to improvement of teaching. One of the "right conditions" for beneficial effects of behavioral feedback is that instructors are sufficiently motivated and open-minded about teaching to put effort into reading and thinking about the TBI feedback provided. This condition, in my experience, is more likely to be achieved with graduate students than with regular faculty instructors, and this may be one reason for the particularly clear results of the Murray and Smith study. Maybe, as the saying implies, it is easier to teach new tricks to young dogs than to old dogs?

Behavioral Training

A second way of applying research on low-inference teaching behaviors to improvement of teaching is through the design of faculty development programs that provide intensive training on a limited subset of classroom behaviors known to contribute significantly to instructional outcome measures. As one example of this sort of effort, Murray and Lawrence (1980) assessed the impact of training in speech and drama skills on the classroom teaching of university professors. The rationale of such training is that the same expressive behaviors used by actors to convey meaning on the stage—for example, vocal variation, movement and gesture, facial expression, pausing and eye contact—can be used by teachers to communicate more effectively in the classroom. Given that teacher enthusiasm/expressiveness has consistently been found to correlate highly with student instructional ratings, it was expected that training of expressive teaching behaviors would produce significant improvement in rated teaching effectiveness. The impact of speech and drama training was assessed by a nonequivalent control group, pretest-posttest design in which the participating instructors were 24 full-time faculty members in the Departments of Psychology, Sociology, and Physics at the University of Western Ontario. An experimental group of 12 teachers volunteered (and paid) for a series of 20 two-hour training sessions taught by a professional actor who also worked as a speech and drama instructor. Specific activities in weekly sessions included breathing and voice exercises, reading of monologues, acting out of short scenes from plays, and delivery of videotaped mini-lectures with corrective feedback from the instructor during playback. In all of these activities, participants were encouraged to make full use of expressive communication behaviors. A control group of 12 teachers, matched with the experimental group in terms of academic discipline and years of teaching, were assessed at pretest and posttest stages of the experiment, but received no behavioral training. Student ratings of both specific low-inference teaching behaviors and of overall teaching effectiveness were obtained just prior to (pretest) and immediately following (posttest) the 20-week training program for both experimental and control teachers. Similar overall effectiveness ratings were obtained for both groups in the four courses most recently taught by each teacher prior to the onset of the speech and drama program. As depicted in Figure 3, it was found that neither experimental nor control teachers showed improvement in teaching prior to the advent of the program, but experimental teachers then showed significant gains in student ratings from pretest to posttest, whereas control teachers showed no measurable change during the same time frame. The absence of improvement in experimental teachers prior to program onset suggests that pretest to posttest gains for the experimental group were due to behavioral training per se, rather than to greater motivation to improve in experimental teachers. To investigate further the impact of the program, the various low-inference teaching behaviors assessed at pretest and posttest stages were classified as either "target behaviors" that were expected to change as a result of the program (e.g., speaks expressively, lectures without notes, facial expressions) or "nontarget behaviors" that were not expected to change (e.g., provides sample exam questions, addresses students by name). As may be noted in Figure 3, the experimental group showed significantly larger pretest to posttest gains than the control group for target teaching behaviors.
but not for nontarget behaviors. This result gives more credibility to the conclusion that the speech and drama program, rather than some extraneous variable, was responsible for the improvement in teaching exhibited by experimental teachers. Thus, as is the case with behavioral feedback, it appears that under the right conditions, a training or faculty development program focusing on a limited set of low-inference teaching behaviors can lead to significant improvement in quality of teaching.

In summary, it appears that research on low-inference teaching behaviors can provide a basis for effective programs for improvement of university teaching. Although teaching improvement is a complex, multifaceted, and long-term problem, with many possible models or approaches, there are some important advantages of incorporating low-inference teaching behaviors into teaching improvement programs. For one thing, because such behaviors tend to be very specific and concrete, they are relatively easy for faculty development consultants to define or describe, and relatively easy for faculty members to comprehend and modify. Second, because low-inference teaching behaviors can be viewed as the “leading edge” of teaching, the point of direct contact between teacher and student, it would seem that they are more likely to have an impact on student development than are more “abstract” or “cognitive” teacher characteristics such as goals, attitudes, knowledge, or planning. As a case in point, it does no good to be well-prepared for teaching or to be intensely enthusiastic about the subject matter in your own mind unless this preparation or enthusiasm is communicated by way of specific teaching behaviors that are observable to students in the classroom. Third, unlike some other models and approaches to teaching improvement, there is systematic research evidence that low-inference teaching behaviors are in fact causally related to student ratings as well as student learning, and can in fact contribute significantly to improvement of teaching.

CONCLUSIONS

Recent research on low-inference teaching behaviors and teaching effectiveness in higher education supports the following general conclusions:

1. There are specific, observable classroom teaching behaviors that account for a large proportion of the variance in student ratings of college and university teaching.
2. In addition to student ratings, low-inference teaching behaviors are related to a wide range of other measures of teaching effectiveness, including student learning and student motivation for further learning.
3. Low-inference teaching behaviors appear to be “causes” rather than simply “correlates” of various measures of teaching effectiveness, suggesting that incorporation of low-inference
behaviors into teaching improvement programs is likely to lead to actual improvement.

4. Very little is known about the cognitive or affective processes underlying the impact of low-inference teaching behaviors, but preliminary research suggests that teacher enthusiasm may affect student learning by way of memory encoding and cognitive structuring mechanisms. More research is desperately needed in this area.

5. Under the right conditions, feedback and training programs based on low inference teaching behaviors can lead to significant improvement in teaching.

6. Although low-inference teaching behaviors have been found to differ in frequency of occurrence across different contexts (academic disciplines, teacher genders, and class sizes), research conducted to date suggests that the contribution of specific teaching behaviors to overall teaching effectiveness tends to be consistent across different contexts or situations.

REFERENCES


RESEARCH ON LOW-INFERENCE TEACHING BEHAVIORS: AN UPDATE

Harry G. Murray
University of Western Ontario
murray@uwo.ca

Abstract

This chapter provides an update to earlier reviews of research by the author (Murray, 1991, 2001) on the contribution of low-inference classroom teaching behaviors to university teaching effectiveness. Research reviewed here supports the following conclusions: (1) the covariance structure of low-inference student ratings resembles the covariance of actual teaching behaviors more closely than does the covariance structure of high-inference student ratings, indicating that low-inference ratings are higher in factorial validity; (2) low-inference teacher clarity behaviors contribute positively to both student achievement and student motivation, but contrary to expectation, these effects do not depend on student anxiety; and (3) high school student ratings of low-inference teaching behaviors show reasonable levels of rater accuracy and have potential as a source of formative feedback for high school teachers.

Key Words: Teacher characteristics, low-inference teaching behaviors, student instructional ratings

The goal of research on teacher effectiveness in higher education is to identify characteristics of teachers that contribute significantly to student cognitive and attitudinal outcomes. It is assumed that knowledge of factors contributing to effective teaching will lead both to a better theoretical understanding of teaching and to the development of improved programs of faculty selection, evaluation, and development.

My own research on teacher effectiveness has focused mainly on specific, concrete “low-inference” classroom teaching behaviors, such as “signals the transition from one topic to the next”, “provides a preliminary outline of the lecture”, and “addresses individual students by name”, that can be can be recorded objectively, with little or no judgement or inference, on the part of a classroom observer. High-inference teacher characteristics, on the other hand, are global, abstract traits such as “clarity” and “satisfaction”, the assessment of which requires more inference or subjective judgement.

Although knowledge of both low-inference and high-inference characteristics is needed for a full understanding of teaching effectiveness, there are some definite advantages in focusing on low-inference teaching behaviors. Such behaviors are relatively easy to manipulate or record for research purposes, and feedback on low-inference behaviors is useful for faculty development purposes because it provides specific, concrete information on what needs to be done to improve teaching. Also, low-inference classroom teaching behaviors represent the “leading edge” of teaching, the point of direct contact between teacher and student, and thus would appear to have the most direct impact on student learning and development.

Previous research by myself and others, reviewed in earlier volumes of Higher Education: Handbook of Theory and Research (Murray, 1991, 2001), showed that there are clear relationships between low-inference classroom teaching behaviors and several outcome measures, including student ratings of instruction, student motivation, and student learning. This research included classroom observation studies, in which trained observers visited class sessions to record the frequency of occurrence of specific teaching behaviors, which were then related to student outcome measures, and experimental studies in which low-inference teaching behaviors were experimentally manipulated in a laboratory setting, with other variables controlled, to demonstrate cause-effect relations with outcome variables.

For purposes of classroom observation studies, I developed an instrument known as the Teacher Behaviors Inventory (TBI), which is used by classroom observers to estimate the frequency of occurrence of each of 50 low-inference teaching behaviors on a 5-point rating scale ranging from 1 (almost never) to 5 (almost always). There are actually two versions of the Teacher Behaviors Inventory: a Research Version, intended for research purposes as defined above, and a Feedback Version intended solely to provide diagnostic or formative feedback to instructors seeking to improve their teaching. The TBI Feedback Version consists of the same 50 low-inference behavioral items as the Research Version, but instead of rating frequency of occurrence per se, students or other observers use a bipolar 5-point rating scale to indicate whether, for purposes of teaching improvement, the instructor needs to increase (rating of +1 or +2), decrease (rating of −1 or −2), or make no change (rating of 0) the frequency of occurrence of
each low-inference teaching behavior. Both versions of the TBI are uncopyrighted, and may be viewed or copied online at the following website: www.ssc.uwo.ca/psychology/faculty/murray.

The remainder of this chapter provides an update of recent research on low-inference classroom teaching behaviors. Most of this research was conducted by graduate students I have worked with at the University of Western Ontario. Three different questions are addressed in this research: (1) the factorial validity of student ratings of low-inference teaching behaviors, (2) the interaction of low-inference teacher clarity behaviors with student anxiety, and (3) the validity and utility of high school student ratings of low-inference teaching behaviors.

FACTORIAL VALIDITY OF LOW-INFERENCE STUDENT RATINGS

This research was reported in a masters thesis by Robert Renaud (1996), and subsequently published by Renaud and Murray (2003). The goal of the research was to compare the factorial validity of student ratings of global, high-inference teacher characteristics vs. ratings of specific, low-inference teaching behaviors. Factorial validity is defined as the extent to which the correlation found among rated components of performance resembles the actual or true correlation of these components.

Although Cohen (1981) and Feldman (1989) have shown that student ratings of global, high-inference teacher characteristics such as clarity, rapport, and organization, are "valid" in the sense that they correlate significantly with student achievement on common final exams, it is possible that global components of teaching effectiveness are actually very low in factorial validity. Cronbach (1958) argues that people have a tendency to rate performance according to "what is thought to go with what", rather than "what actually goes with what", and this over-reliance on conceptual associations explains how traits can be rated as correlated when in reality they correlate little or not at all. For example, instructors who are well liked by students may be rated highly on both "Rapport" and "Organization", even though there is no actual or true tendency for teachers with good rapport to also be highly organized.

The recommended procedure for examining the impact of conceptual associations on performance ratings is to compare three correlation matrices: one based on ratings of component factors in performance, one based on conceptual associations among the same components, derived by having raters determine in their own minds how strongly each dimension is associated with each other dimension, and one based on direct observation of actual behaviors underlying each component of performance. Factorial validity is indicated if the rated component matrix shows a strong correlation with the actual behavior matrix, and a weaker correlation with the conceptual association matrix.

Although several previous studies (e.g., Cadwell & Jenkins, 1985; Whitley & Doyle, 1976) have reported that student ratings of teaching are strongly influenced by conceptual associations, an important limitation of these studies is that they failed to compare component ratings and similarity judgements to actual teaching behaviors. The degree to which ratings are influenced by conceptual associations is difficult to confirm in the absence of direct observation of teaching behaviors. Another limitation is that previous studies focused solely on student ratings of global, high-inference teacher characteristics, whereas higher levels of factorial validity might be found for student ratings of specific, low-inference teaching behaviors that require little subjective judgement on the part of the rater.

Data were obtained from 32 instructors videotaped during two 1-hour class periods, and from three separate groups of graduate students: the rating group, consisting of 622 students registered in the 32 classes taught by the participating instructors; the similarity judgement group, consisting of 43 students who provided pairwise similarity judgements of the same low- and high-inference teacher characteristics rated by the rating group; and the observer group, consisting of 256 students who recorded the frequency of occurrence of actual teaching behaviors from videotape of classes taught by the 32 instructors.

Four different teacher rating forms were used in this study. The first was an abbreviated version of the Teacher Behaviors Inventory, consisting of 16 low-inference classroom teaching behaviors representing 8 dimensions of teaching, with each behaviour rated on a 3-point frequency-of-occurrence scale (1 = "almost never", 3 = "almost always"). The second was the Teacher Rating Form (TRF), a newly developed 9-item form consisting of 8 high-inference items corresponding to the 8 dimensions represented on the TBI, plus a ninth item assessing overall teaching effectiveness. Each TRF item was rated on a 5-point scale ranging from 1 = "poor" to 5 = "excellent". The third rating form, used for the conceptual association task, included all
possible pairs of items from the 16-item TBI, and all possible pairs of items from the 8-item TRF (excluding overall effectiveness). Subjects rated the degree of conceptual similarity of each pair of items on a scale ranging from $-100$ meaning "completely opposite" to 0 meaning "no similarity at all" to $+100$ meaning "completely similar." Finally, subjects in the observer group used a behaviour observation form to record each of the 16 TBI behaviors as either present or absent in each of 24 sequential 5-minute time segments, such that the recorded frequency of a behavior could range from 0 to 24 in 120 minutes of videotape.

Preliminary data analyses showed that the mean inter-rater reliability (intraclass correlation) of student ratings of the 16 low-inference TBI items was .82; whereas the mean inter-rater reliability of student ratings of the 8 high-inference TRF items was .80; and the mean inter-observer reliability of counts of actual teaching behaviours was .88. Thus, reliability of measurement was generally high, particularly for direct observation or rating of specific classroom teaching behaviours.

Covariance matrices for low- and high-inference student ratings of teaching were obtained by intercorrelating all TBI or TRF items using instructor mean scores as the unit of analysis (N = 32). Corresponding matrices representing similarity judgements for TBI and TRF items were obtained by using the mean judgement of similarity for each pair of items for 43 judges. Thus, there were three 16 x 16 matrices representing the TBI items across each of the three tasks, and three 8 x 8 matrices representing the TRF items across the same three tasks.

To assess similarity of covariance structures across the three matrices for TBI and TRF items, each matrix was correlated with each other matrix by rearranging the values in the bottom half of each matrix into a column vector, and intercorrelating the three vectors. The top panel of Figure 1 shows correlations among the three matrices for low-inference TBI items. The pattern of covariance among teacher ratings corresponded closely to the pattern of similarity judgements, $r = .81$, $p < .01$, and also resembled the pattern of correlation among actual behaviors, $r = .54$, $p < .01$. Furthermore, to a slightly lesser degree, the pattern of correlation among actual behaviors was significantly related to the similarity judgements matrix, $r = .42$, $p < .05$. The bottom panel of Figure 1 shows correlations among the three matrices for high-inference TRF items. These data suggest a somewhat different picture. Unlike the TBI matrices, the only significant relation among TRF matrices occurred between the ratings task and the similarity judgement task, $r = .53$, $p < .01$.

The results of this study are consistent with previous evidence (e.g., Murray, 1983) that student ratings of low-inference teaching behaviors are slightly higher in intrarater reliability than student ratings of high-inference teacher characteristics. Secondly, the present findings suggest that ratings of low-inference teaching behaviors are higher in factorial validity than high-inference ratings, in that the covariance pattern for low-inference ratings related not only to students' mental conceptions of which behaviors go together, but to actual covariance of directly observed classroom teaching behaviors, whereas the covariance structure of high-inference ratings related only to student' conceptual associations.

Thus, when student ratings are used by promotion and tenure committees to judge teaching, or by faculty development specialists to improve teaching, particularly when components of teaching are used in a profile format to indicate areas of relative strength and weakness, low-inference student ratings may provide a more accurate and useful measure of relative performance on various components of teaching effectiveness than high-inference student ratings.
INTERACTION OF TEACHER CLARITY AND STUDENT ANXIETY

This research was part of a doctoral dissertation completed by Susan Rodger (2001) and since accepted for publication (Rodger, Murray & Cummings, 2007). A laboratory experiment was carried out to examine a possibility that teacher clarity and student test anxiety enter into an aptitude-treatment interaction (Cronbach & Snow, 1977), whereby the effect of teacher clarity is greater for high than for low anxiety students. The predicted aptitude-treatment interaction (ATI) was examined in relation to two different outcome measures, namely student achievement and student motivation for further learning.

Teacher clarity has consistently been found to positively influence student outcomes in previous research (Murray, 2001). In the present study, teacher clarity was a manipulated treatment variable, defined operationally by the following set of nine low-inference teaching behaviours: puts outline of lecture on projection screen, uses concrete examples, uses multiple examples of each concept, repeats difficult ideas, suggests practical applications, stresses important points, signals transitions between topics, summarizes periodically, and highlights similarities and differences between concepts.

Test anxiety has consistently been found to have a negative effect on student academic achievement and motivation (Hembree, 1988), possibly because it is associated with deficits in information processing. It was hypothesized that the positive effect of teacher clarity on student achievement and motivation would be larger for students high in test anxiety in the present study. One reason for expecting such an ATI is that the low-inference teaching behaviours underlying teacher clarity can be assumed to assist students in encoding and organizing new information, and highly anxious students with presumed deficits in information processing would be expected to benefit the most from high clarity teaching.

The research participants, 120 first-year psychology students, were randomly assigned to watch either a Low Clarity or High Clarity videotaped lecture on the topic of Memory and Amnesia and then read an assigned paper on the same topic. The two videotaped lectures were identical in content coverage and in length, the only difference being that the High Clarity lecture incorporated the 9 low-inference teaching behaviours listed above, whereas the Low Clarity lecture incorporated “filler material” such as historical facts and superfluous details. Test anxiety was measured by Spielberger’s (1970) Test Anxiety Inventory (TAI), and the 60 students at each level of teacher clarity were divided at the median TAI score to create Low Anxiety and High Anxiety conditions within a 2x2 factorial design. Student achievement was assessed by a multiple-choice and short-answer test based on the videotaped lecture and assigned reading, whereas student motivation was measured by Pintrich et al.’s (1991) 17-item Motivated Strategies for Learning Questionnaire.

Students watched either the Low Clarity or High Clarity lecture in the first experimental session, then returned one week later after completing the assigned reading to take the achievement test covering both the lecture and reading. It was assumed that this design provided a laboratory simulation that replicated as closely as possible the conditions of real-world teaching and learning, including a lecture and assigned reading followed by a test.

The major results of the study are summarized in Figure 2. As may be noted in the top panel of the figure, both of the expected main effects were obtained for the student achievement measure, in

Figure 2: Effects of teacher clarity and student anxiety on student achievement (top panel) and student motivation (bottom panel).
Thus, the present study shows that teacher clarity is a true causal antecedent, rather than merely a correlate, of student learning, and this increases our confidence that efforts by teachers to increase their use of low-inference “clarity” behaviors will actually pay off in terms of improvement in student achievement and motivation. The underlying process or mechanism whereby teacher clarity facilitates student learning is uncertain, but may relate to the role of underlying low-inference teaching behaviors in structuring information in short-term memory, thus leading to more meaningful encoding of information in long-term memory.

Another important contribution of this study is the demonstration that teacher clarity benefits student motivation as well as student learning. The theoretical explanation for this effect is also open to speculation, but it seems possible that teachers who use multiple examples, repeat difficult ideas, and provide a lecture outline create a sense of self-efficacy or personal control in students, thereby improving motivation to learn.

HIGH SCHOOL STUDENT RATINGS OF LOW-INFEERENCE TEACHING BEHAVIORS

Two studies reported in a doctoral dissertation by Kristin Anglin-Bolting (2005) were designed to investigate the validity and utility of high school student ratings of low-inference teaching behaviors. The goal of the first study was to compare low-inference ratings made by high school vs. university students in terms of Cronbach’s (1955) four components of rating accuracy. As explained below, Cronbach’s model assesses rating accuracy in terms of the extent to which student ratings of teaching are similar to those of expert raters.

The first component of accuracy in Cronbach’s model, elevation (EL) is the degree to which a rater, on average, rates performance either too leniently or too severely relative to an expert rater. The second component, differential elevation (DE), is defined as the rater’s ability, relative to that of an expert, to differentiate the performance levels of different individuals for summative evaluation, while controlling for overall rating elevation. Accuracy (SA), the third component, represents the rater’s ability to distinguish between different performance items averaged across individuals, with overall elevation controlled. The fourth component of rating accuracy, differential accuracy (DA), measures the accuracy of identifying an
individual's performance profile or pattern of strengths and weaknesses for purposes of formative evaluation.

There were three groups of raters in this study: experts, undergraduate students, and high school students, all of whom watched three videotaped simulations of high school teaching and then rated low-inference teaching behaviours. The sample of 16 expert raters included elementary school teachers, as well as graduate students and faculty members in educational psychology. The undergraduate students were 61 introductory psychology students, and the high school students were 164 volunteers, aged 14 to 19, representing all five grade levels in a local high school. Each of the three videotaped lessons was presented by a different experienced teacher, and each dealt with a separate topic relating to economic systems and the economy of China. The first videotape was used for practice and feedback purposes only, whereas the second and third videotapes were used to obtain rating accuracy estimates. Low-inference teaching behaviors in the three videos were evaluated using a subset of 15 items measuring three categories of teaching (clarity, enthusiasm, and voice quality) from the High School Teacher Behaviours Inventory (HSTBI). The HSTBI is a modified version of the Teacher Behaviours Inventory consisting of 58 items relevant to teaching at the high school level, each rated on a 5-point frequency-of-occurrence scale: Almost Never (1), Rarely (2), Sometimes (3), Often (4), and Almost Always (5).

High school and university students received rater training before rating the second and third videotapes. This consisted of definition of performance dimensions, examples of behaviors associated with various levels of performance, and practice in rating the first videotape followed by feedback relating their ratings to those of the expert raters. Rating accuracy scores for each of Cronbach's four components of accuracy on the second and third videotapes were calculated separately for each student and each teaching behaviour by comparing student ratings to mean expert ratings of the same behaviors.

Table 1 shows means and standard deviations of Cronbach rating accuracy scores averaged across all 15 HSTBI items for high school and undergraduate students. In general, rating accuracy was better (i.e., closer to zero) for undergraduate students than for high school students. Although high school students were slightly more accurate in terms of the elevation (EL) component of accuracy, undergraduate students were more accurate in terms of the other three Cronbach components, namely differential elevation (DE), stereotype accuracy (SA), and differential accuracy (DA). Across all four accuracy components, the multivariate main effect for student group was significant in favor of university students. Further analyses of differences between university students and specific age groups at the high school level showed that ratings by 17, 18, and 19 year old high school students were significantly less accurate in terms of differential elevation (DE) than were ratings by undergraduate students, whereas 16 year old high school students gave significantly less accurate ratings in terms of stereotype accuracy (SA), and 15 year old high school students gave significantly less accurate ratings in terms of differential accuracy (DA). No other differences across age groups were significant.

The results of this study suggest that although there were significant differences between high school and university students for three of the four components of accuracy, these differences were not large or consistent or widespread. For example, ratings from 14 year old high school students did not differ significantly from undergraduate student ratings on any component of accuracy, and the oldest high school students were the least accurate in terms of distinguishing different performances across teachers (DE). In addition, high school students of varying ages did not differ significantly from each other. It appears that high school students are capable of rating teaching performance at close to the same level of accuracy as university students. Worrell and Kuterback (2001) reached a similar conclusion in a recent study where high school student ratings of low-inference teaching variables showed the same factor structure and same relationship to overall teaching effectiveness as has been reported at the university level.
Anglin-Bodrug (2005) conducted a second study to evaluate the effect of feedback from high school student ratings of low-inference teaching behaviors on the subsequent teaching performance of preservice teachers. High school students were asked to provide low-inference ratings of preservice teachers after two separate practice teaching placements, once at Time 1 and then again at Time 2. Preservice teachers were randomly assigned to either a feedback group or to a "waiting-list" control group. Those in the feedback group received the results of student ratings immediately after completing their Time 1 practice teaching placement, while the control group received feedback only after the study was completed (i.e., after Time 2). It was expected that the feedback group would show greater improvement in performance ratings from Time 1 to Time 2 than the control group.

The participants in this study were 93 preservice teachers who volunteered to participate, and were asked to select classes taught in their next two practice teaching placements wherein student ratings of teaching would be solicited. Student ratings of low-inference teaching behaviors were obtained via a subset of 28 items from the High School Teacher Behaviors Inventory, covering 7 dimensions of classroom teaching. The number of high school students providing HSTBI ratings was 901 at Time 1 and 946 at Time 2.

Preliminary analysis of data showed that, despite random assignment of subjects to conditions, self-ratings of teaching were significantly higher for the Control Group than for the Feedback Group prior to the onset of differential treatment. Figure 3 shows mean student ratings averaged across all 28 HSTBI items for Feedback and Control groups at Time 1 and Time 2, with self-ratings statistically controlled. It may be noted that mean student ratings were similar for Feedback and Control groups at Time 1, whereas at Time 2 there was an increase in student ratings for the Feedback Group, but no change in ratings for the Control Group. This pattern of results is consistent with expectation, but statistical analysis failed to show the expected Groups × Times interaction effect. This result is surprising for several reasons: (1) in terms of effect size, the present results look very much like those reported by Cohen (1980) for similar feedback studies at the university level; (2) teacher self-reports indicated that student feedback was helpful and beneficial; and (3) significant effects of high school student feedback on low-inference teaching behaviors were reported in a parallel study by Smits (2002). It is possible that the non-significant feedback effect in the present study was due to high within-group variability in student ratings and lack of statistical power.

In summary, the results of the two studies reported above suggest that high school student ratings of low-inference teaching behaviors could potentially be used for both summative and formative evaluation of teaching. The present research indicates that high school students of all age levels, like their university counterparts, can provide accurate low-inference teacher ratings, and these ratings could potentially be effective as a source of feedback to improve teacher performance.

CONCLUSIONS

The results of the three studies reviewed above confirm and extend previous evidence (Murray, 1991, 2001) demonstrating the value of focusing on specific, low-inference classroom teaching behaviors in research on teacher effectiveness, in evaluation of teaching, and in faculty development programs in higher education. The results of Study 1 suggest that student ratings of low-inference teaching behaviors are less influenced by implicit assumptions and conceptual associations than ratings of high-inference teacher characteristics, and thus are potentially useful for both summative and formative evaluation of teaching. Study 2 demonstrated a strong cause-effect relationship between teacher clarity and both student learning and student motivation. The fact that teacher clarity effects were similar for low and high anxiety students was contrary to prediction, but
consistent with previous evidence (Murray, 2002) that effects of low-inference teaching behaviors are highly stable or robust across different contexts. The results of Study 3 suggest that high school students are capable of providing accurate ratings of low-inference teaching behaviours, and such ratings could potentially be used for both summative and formative purposes in high schools.

REFERENCES


7. TEACHERS' NONVERBAL BEHAVIOR AND ITS EFFECTS ON STUDENTS

Elisha Babad
Hebrew University of Jerusalem
elisha@vms.huji.ac.il

Abstract
This chapter presents the area of nonverbal (NV) behavior as it relates to teacher-student interaction, particularly in higher education. The first part covers research topics in NV psychology, the repertoire of NV behaviors, and copies of NV research pertinent to teaching. Microteaching is then discussed as a major application in teacher training involving NV behavior. The central part focuses on instructors' NV behavior and its effects on students. The "teacher enthusiasm" and "teacher immediacy" conceptualizations and research literatures are then discussed, wondering about their integrated isolation from each other, because both deal with the very same phenomenon of the contribution of instructors' NV enthusiasm to their teaching quality. Research on specifically-measured instructors' NV behaviors (opposed to global NV conduct as perceived by students) is then presented, demonstrating how thin slices (10 seconds) of teachers' NV behavior can predict student evaluations, and illuminating the NV profile of effective instructors.

Key Words: Nonverbal (NV) behavior; NV psychology; teacher enthusiasm; teacher immediacy; Doctor Fox research; instructors' NV behavior; teachers' NV behavior; thin slices research; students' ratings of teaching (SRT); students' evaluations of teaching (SET); microteaching; teacher-student interaction; effective teaching; micro-analysis; NV profile of effective instructors

INTRODUCTION: IMPORTANCE OF NONVERBAL BEHAVIOR

We all live today in the era of the visual, of the nonverbal. People are continuously and excessively exposed to television, cinema and theater, and these media transmit a multitude of types and bits of information in nonverbal (NV) channels. From a young age children learn...